AFRL-SR-AR-TR-09-0289

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*: 30-07-2009

2. REPORT TYPE: Final Report

3. DATES COVERED *(From - To)*: May 1, 2007 - April 30, 2009

4. TITLE AND SUBTITLE
Choosing among Causal Agents in a Dynamic Environment

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0429

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Young, Michael E.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Southern Illinois University
Mailcode 4709, Woody Hall C-206
Carbondale, IL 62901

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Southern Illinois University
Office for Research Development and Administration
Mailcode 4709, Woody Hall C-206
Carbondale, IL 62901

10. SPONSOR/MONITOR'S ACRONYM(S)
ORDA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approval for public release; distribution is unlimited

20090914179

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Participants in a video game environment were required to make a series of decisions in which they must identify which of three targets was causing a distal explosion. The potential targets were firing weapons which could produce an explosion after a constant or variable delay, a delay that was filled or unfilled with an auditory event, and may have produced explosions probabilistically. Results: delays had profound effects on accuracy and decision latencies, decreasing weapon effectiveness from 100% to 50% had little effect on accuracy and modest effects on latencies (men only), filling a delay helped under very limited conditions, and varying the delay actually improved performance for longer (average) delays. Furthermore, men's decision accuracy was higher but not when prior video game experience was controlled for. In contrast, women observed their targets for much longer before making a choice regardless of prior experience. The results disconfirmed the proposed forward inference model and instead supported the operation of a backward inference model of causal choice.

15. SUBJECT TERMS
causality; decision making; stress; delay; individual differences

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU		Michael E. Young

19b. TELEPHONE NUMBER *(Include area code)*: 618/453-3567

FINAL REPORT

AFOSR Program: Mathematical Modeling of Cognition and Decision

FA9550-07-1-0429

Choosing among Causal Agents in a Dynamic Environment

Michael E. Young

Email: meyoung@siu.edu

Research Objectives

The two-year AFOSR-funded project involved the study of causal perception, learning, and judgment when the pertinent events involved moving, interacting objects. The principal objective was to determine the extent to which a theory that was a product of controlled experiments involving a single causal candidate, only one visual perspective, and minimal experienced stress, would pertain to a video game environment that involved multiple causal candidates, many visual perspectives, and varying degrees of stress. The theoretical basis for these studies is that causal judgments rely on the subjective certainty regarding *whether* an effect will occur and *when* it will occur. The experiments were designed to address the following specific aims:

- Evaluate the relative efficacy of various delay fillers and whether the efficacy of these interventions can be forecast by a proposed *temporal contingency model*, and
- Assess the model's predictions relating to the integration of causal cues regarding contingency, endogenous temporal variability, and exogenous temporal variability, and
- Measure the ways that a key independent variable in real-world causal decisions, *stress*, affects the utilization of causal cues.

The Psychological Study of Causality

Historically, the judgment of causality has been assumed to heavily rest on the notion of contingency – to what extent does the likelihood of the effect of interest depend on the occurrence of each candidate cause, P(E | C1) vs. P(E | C2), etc.? This historical emphasis began with the formal testing of a specific predictor of human causal choice, delta-P or ΔP, that is based on the difference between the effect's likelihood given the occurrence of a candidate cause versus the effect's likelihood when the candidate cause does not occur: $\Delta P = P(E \mid C) - P(E \mid \text{not } C)$ (Allan, 1980; Kao & Wasserman, 1993; Shanks, 1995). The emphasis on the relationship between conditional probabilities continued with Cheng's (1997) work on causal power and also underlies modern Bayesian perspectives on causal judgment (e.g., Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Tenenbaum, Griffiths, & Kemp, 2006). All of these accounts

emphasize the forward inferential aspect of causality – the likelihood that the effect will follow given that one or more candidate causes have occurred.

In our AFOSR-funded project, we evaluated a temporal extension of forward inference by examining the temporal density profile of the effect following each of the candidate causes and predicted that people's choice among causal alternatives would be a byproduct of the relative peaks of each of these profiles. In other words, participants needed to identify which among the causal candidates produced the greatest temporal certainty regarding the effect. This approach hypothesized that causal decisions are based on both uncertainty regarding *whether* the effect would occur and *when* it would occur. Unfortunately, the data did not bear out our theoretical notions of the importance of causal candidates' temporal density profiles. This failure required a rethinking of the variables that were affecting people's decisions in the complex dynamic environments that our participants faced.

As we investigated the behavioral patterns actually observed, we reached the conclusion that participants were looking backward, not forward, in their judgments of causality. More formally, participants were less concerned with the $P(E \mid C)$ for each candidate cause and more with the $P(C_{contiguous} \mid E)$ for each candidate cause where $C_{contiguous}$ designates the candidate cause most recently experienced at the time of the effect. In everyday terms, participants' choices were strongly determined by which of the candidate causes was more likely to have been the most proximal to the effect. Proximity can only be judged in retrospect – once the effect has occurred, the participant must look backward to identify which of the candidate causes occurred most recently.

Although there is evidence that people can reach rational conclusions about the Bayesian nature of causal relations when given sufficient time, simple environments, and the ability to intervene and thus test hypotheses regarding each cause and its interactions with other events (Gopnik et al., 2004; Waldmann & Hagmayer, 2005), the story was quite different for our participants. Within the context of our video game there were at least three alternative causes occurring at random times relative to one another, possible delays between the cause and its effect, no ability to intervene in an attempt to control when the cause would occur (the opponents shoot when they want, not when the participant wants), and choices that were sometimes required to be made under time pressure. This greater complexity appears to have produced a narrower time horizon for decisions such that choice and latencies were primarily driven by a temporally extended computation related to the $P(C \mid E)$. Specifically, once the effect occurred (an explosion), participants examined the recent occurrences of the candidate causes (weapons' discharges) and identified which one was most proximal to the explosion. If, after repeated observation, one of the weapons produced more proximal discharges than the others, then that weapon was chosen as the source of the explosion.

Prior results

In our 2-year AFOSR-funded project (2007-2009), we evaluated an earlier model of the attribution of causality in complex dynamic environments. The *temporal contingency model* hypothesized that "a candidate cause will be more likely to be considered as a cause to the extent that it predicts both *whether* and *when* an effect of interest will occur." To evaluate this hypothesis and various specific mathematical models derived from the more general model, we conducted a series of experiments examining people's behavior (choice accuracy and latency)

when choosing among three alternative causes, one of which was the true cause of a distal explosion. Characteristics of the cause that were varied included delays (0.5, 1.0, or 2.0 s), the presence of auditory delay fillers designed to improve temporal predictability, delay variability (constant cause-effect delays or low or moderate variation in the cause-effect delays), and likelihoods (50%, 75%, and 100%). Some of these factors varied across groups of opponents within a game level and some varied across game levels; any single experiment only varied two of these causal characteristics. Additionally, we examined the effect of time pressure on choice accuracy and latency when the opponent's weapons operated with a delay and with probabilistic outcomes. Individual differences analyses were conducted to determine if sex or prior video game experience would have independent effects on choice (cf. Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Feng, Spence, & Pratt, 2007; Green & Bavelier, 2003, 2006, 2007; Young & Nguyen, 2009).

There were seven key findings across our studies:

- Changes in the delay between a weapon's discharge and the explosion had profound effects on accuracy and latency. A delay as short as 2.0 s produced near chance or below chance accuracy.
    - When we changed the overall firing rate to determine whether there was something special about 2.0 s or whether the key factor was the ratio of the delay to the average interfiring interval (IFI), we discovered that participants modulated their latencies to compensate for changes in the IFI. When the IFI was shorter (which should make delays more problematic), participants were able to maintain their accuracy by increasing their latency.

- Adding variability to a delay had little impact when the delays were short but *improved* accuracy when the delays were 2.0 s. This single result was especially problematic for the proposed temporal contingency model and necessitated the development of an alternative theory of causal decisions within these complex environments.

- Filling the delay had little impact except when the average delay within a game level was held constant (accuracy improved by 10% when the delay was filled).

- Varying outcome likelihood in the 50% to 100% range had little discernible effect on choice accuracy although lower likelihoods produced modest increases in latencies for men.

- Men showed higher accuracies (typically about 10% higher) and shorter latencies (10-20 s shorter) in all experiments. The difference in accuracy was often explained by men's greater self-reported experience playing certain types of video games (usually first-person shooter and strategy games). In contrast, women's latencies remained about 10 s longer than men's even after prior video game experience was partialled out.

- Time pressure had a modest and inconsistent detrimental effect on men's accuracies but had no effect on women's behavior (accuracy or latency).

   - The failure to observe a significant effect of time pressure (specifically, making the decision while under enemy fire) prompted us to try a different time pressure manipulation in a study recently completed.

- Increasing the number of alternatives from two to four produced longer initial choice latencies but no effect on discriminability (assessed using d' to control for the varying levels of chance performance).

- Finally, none of our variables had much of an effect on women's choice latencies – women generally waited the same (long) time regardless of the delay, delay variability, presence of a filler, likelihood, or time pressure.

The interesting, albeit complicating, factor throughout all of the experiments was the ceding of some of the control to the participant. A game player determined how long to observe a series of weapons discharges and explosions before making a decision, what angle and distance to view from, how rapidly to fire their own weapon, and how to navigate the environment in order to move from one decision point to the next. This method was chosen in order to increase the external validity of the experimental results.

Our original AFOSR proposal had the stated goal to submit four manuscripts based on the funded studies. The first has been published in *Learning and Motivation* (Young & Nguyen, 2009), a second has been reviewed at the *Journal of Behavioral Decision Making* (Young, Sutherland, Nguyen, & Cole, in revision) and a revision has been requested, a third will describe the emerging theoretical framework that arose from this research and will appear in an invited book chapter (Young, in preparation for *Handbook of Comparative Cognition*), and a fourth will describe the results recently collected. The AFOSR-funded work has also produced five presentations in major conferences including the annual meetings of the Southeastern Association for Behavior Analysis, the Society for Judgment and Decision Making, and the American Psychological Association (an invited address to be given in August, 2009) as well as invited talks at regional universities (Southern Illinois University at Edwardsville, Western Kentucky University, and Drake University [in October, 2009]). Related work on continuous causation (in which the causes and effects vary in degree) is being conducted with the support of NSF funding.

Individual Differences. Throughout our experiments, we have investigated individual differences in performance as a function of sex, amount of prior video game experience, and type of prior video game experience. Across every one of our experiments, women have shown somewhat lower accuracies and much longer decision latencies. When we statistically controlled for the effects of prior video game experience (both amount and type), the accuracy differences disappeared but the latency differences were unaffected. Amount of prior video experience improves choice accuracy but not latencies, although these effects were strongest among those who we called "gamers" – participants who more often played first-person-shooter games, combat and fighting games, and real-time strategy games (90% of these players were men).

Our manipulated variables (delay, delay variability, outcome likelihood, fillers, stress) often affected decision latency, but mostly for men; women tended to show similarly slow latencies regardless of the experimental conditions. Women's longer latencies could be due to more deliberate decision making, more environmental exploration, slower rates of firing, slower

movement in the game, or difficulty navigating the environment, *inter alia*. We are exploring these issues in subsequent research (see below) but had already determined that firing rates do not differ.

Time Pressure. In the second year of the project, we began investigating the effect of time pressure on participants' performance in the decision environment. Given that there has been no prior work on the effects of this variable on causal decision making, we were shooting in the dark regarding the appropriate amount and type of time pressure. We opted to use the proposed method of placing the player's avatar under fire from snipers in the mountains. Our first pilots revealed that some players tried to hunt down the snipers before fulfilling the requirements of the game. Given that this subverted our goals, we redesigned the game to hide the snipers in such a way that they could not be found (and players were informed of such). The subsequent study involved snipers who fired relatively often at random intervals but each shot did little damage to the player. Indeed, we discovered that nearly all players could complete a game level without their avatar's death by progressing throughout the game at a normal pace. Under these conditions, men showed occasional modest decreases in accuracy often accompanied by a modest decrease in latency (a speed-accuracy tradeoff); women showed no effects of being fired on (Young, Sutherland, Nguyen, and Cole, in revision).

In a pilot study recently completed, we tried a variation on the time pressure manipulation in which the snipers fired much less often but when they did so, the damage done was much greater. The average amount of damage was the same as that used in our earlier study, but we predicted that the lower rate of firing and greater damage would produce less habituation to being hit and thus a greater impact on behavior. The results were profound: men and women showed substantial decreases in accuracy (76% to 65% for men, 78% to 58% for women) with no effect on latencies. We again found that women showed much longer latencies before choosing their target (13 s longer). Women's avatars died more often in the game ($M = 3.6$ deaths) than men's avatars ($M = 1.2$ deaths) and women tended to leave their avatar stationary more often – women were more likely to move their avatar to the next target region, stop moving (thus leaving themselves vulnerable to sniper fire), and observe the enemy firing patterns for longer periods of time (i.e., were more conservative in their decision making). We will be following up this successful pilot study by including our original manipulation of delay to determine how stress interacts with the effects of delay on decision making.

A New Model of Causal Choice: The Relative Contiguity Model

A significant product of our empirical research was the development of a new model of causal decision making in video game environments. People's choices appear to have been heavily influenced by the experienced relative temporal contiguity of the targets to the explosion (see Fernbach & Sloman, 2009; Perales & Catena, 2006, for a discussion of the importance of local cues to causality). To appreciate the impact of delay on experienced contiguity, Figure 1a shows a theoretical model of the *perceived* experienced delays between the true target's weapon's discharge and the explosion. I assume a gamma distribution of uncertainty in the perceived duration of the delay because this assumption nicely describes people's estimation of durations due to the scalar nature of timing (Allan, 1998). The upper curves (peaking near 0.5 s, 1.0 s, and 2.0 s) represent the experienced delays-to-explosion due to time estimation uncertainty when there are no foils (the gamma scaling parameter was set to 5.0 for these hypothetical

curves). The lower curves represent the perceived experienced delays-to-explosion for the target in the presence of a single foil, but *only when the foil did not precede it.* The lower curves are determined by the following function in Mathematica:

```
PDF[GammaDistribution[delay*scale, 1/scale], x] × (1 - CDF[UniformDistribution[{0, 3}], x])
```

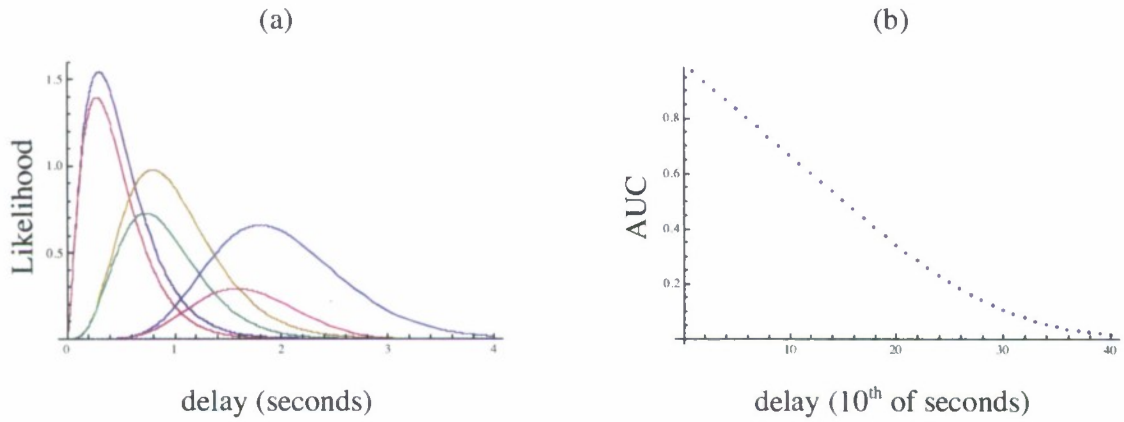

Figure 1

The difference between each pair of curves represents the impact of a foil – it may occur between the time of the true target's discharge and the effect (an explosion) and this likelihood increases as time passes (hence the use of the foil's cumulative distribution function or CDF in the equation). Not surprisingly, a foil has a much larger impact on targets with longer delays because there is more opportunity for a foil to intervene. A summary of the AUC-delay relationship is shown in Figure 1b in which the (foil-adjusted) area under the curve (AUC) is plotted for various delays in the presence of a single foil. The AUC measures the likelihood that a particular candidate cause will be the proximal event to the explosion. The new model supplements this potential predictor of choice accuracy with a second measure: when a candidate cause is the most proximal, what is the experienced delay? Thus, the relative contiguity model predicts that the accuracy and latencies of choices are a function of the AUC for the true target versus that for each foil (i.e., the likelihood that it is the most proximal) *plus* the experienced mean delay (i.e., its degree of proximity).

For the situation shown in Figure 1a, the AUC for all of the 0.5, 1.0, and 2.0 s delays is 1.00 when there are no foils but 0.83, 0.67, and 0.34, respectively, when there is a single foil (although the present theoretical analysis has assumed a gamma scaling parameter of 5.0 to produce a moderate degree of timing uncertainty, the effect of gamma is relatively small except when uncertainty approaches 0). The impact of longer delays on the likelihood of proximity is exacerbated by the additional opportunities for foils to intervene.

Why is average delay also considered as an independent factor in people's choice? This decision is tied to the observed effect of delay variability on accuracy. In a recently published study funded by the AFOSR, we discovered that adding variability to a delay had little impact for short delays but actually improved performance for long delays (Young & Nguyen, 2009). Figure 2 highlights the impact of variability on experienced contiguity for the true cause. For both short and long delays, the AUCs were little changed when variability was added (compare

the narrower curves where there was no variability to the broader curves which included programmed delay variability). In contrast, adding variability to the 2.0 s delays shifts the experienced contiguity curve to the left thus producing a greater proportion of experienced short delays for the true cause.

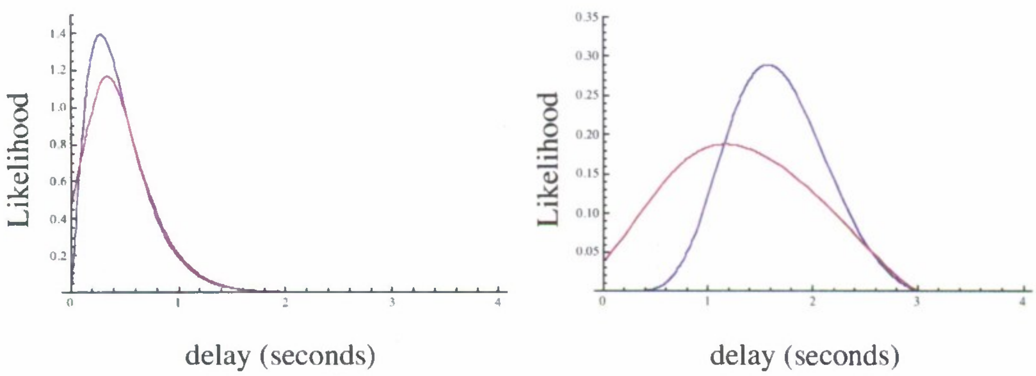

Figure 2

The Effect of Outcome Likelihood. Because our initial experiments only involved one true cause, when this cause produced *probabilistic* outcomes it did not change the experience of a foil intervening between the true cause and the outcome. On those trials when the true cause failed to produce an outcome due to its programmed probability, the foils would likewise not be followed by the explosion. Thus, probabilistic weapons do not create new opportunities for the foil to be more contiguous but rather decrease the overall number of observations of the outcome. This smaller sampling may produce a longer observation time to compensate, a result that we observed in our prior experiments (Young, Sutherland, Nguyen, & Cole, in revision). In the absence of longer observation, the decision would be made after very few observations and this may produce greater emphasis on the mean of the relative contiguity curve rather than its AUC.

To combine the hypothesized mathematical relations discussed, Equation 2 represents a general function that produces a foil-adjusted probability distribution for the true cause's experienced delays (the *rcpdf* or relative contiguity pdf), Equation 3 computes the overall probability of the true cause being the most contiguous to the effect (AUC), and Equation 4 computes the mean experienced delay for the occasions on which the true cause is the most contiguous. The likelihood of a correct response when there is only one foil is computed using a standard logistic relation in Equation 5. Equation 5 allows for the possible interaction between the AUC and mean delay.

$$\text{rcpdf[x\_, delay\_, scale\_, n\_] :=}$$
$$\text{PDF[GammaDistribution[delay*scale, 1/scale], x]} \times \text{(1 - CDF[UniformDistribution[\{0, 3\}], x])}^{n} \qquad (2)$$

$$AUC = \int_0^{\infty} rcpdf[x, delay, scale, n]\,dx \qquad (3)$$

$$\text{Mean} = \text{Expected value of x} = \int_0^\infty x \times rcpdf[x, delay, scale, n]dx \quad (4)$$

$$\text{P(Correct)} = \frac{1}{1 + e^{a \bullet AUC - b \bullet Mean + c \bullet AUC \bullet Mean + d}} \quad (5)$$

The best fitting parameter values of Equation 5 (a, b, c, and d) may vary as a function of the experienced probabilistic efficacy of a weapon, the extant time pressure, individual differences variables, *inter alia*. We will use nonlinear mixed effects modeling (see below) to identify which of these values should be assumed to be constant across individuals or experimental conditions.

Equations 2 through 5 represent the predicted behavior of decision makers in an ideal theoretical world with unlimited sampling. Collectively, they rely on the accurate assessment of the rcpdf. Individual participants, however, will have acquired limited sampling of the rcpdf. Thus, a separate analysis will be performed based on the actual experienced explosion delays for each candidate cause by individual participants. AUC (Equation 3) will be substituted with the relative frequency with which each target is more proximal to the effect, the Mean (Equation 4) will be substituted with the actual mean of these delays, and these experienced values will be used in Equation 5 in order to predict participant choice accuracy.

A proposal involving the evaluation of this new model is under review at AFOSR.

Relevant Publications

Young, M.E. (in preparation). Cues to causality in complex dynamic environments. To appear in *The Handbook of Comparative Cognition*, E.A. Wasserman and T.R. Zentall (Eds.), Oxford University Press.

Young, M. E., Sutherland, S., Nguyen, N., & Cole, J. (in revision). Waiting to decide helps in the face of *whether* uncertainty but not *when* uncertainty. *Journal of Behavioral Decision Making*.

Young, M. E., & Sutherland, S. (in press). The spatiotemporal distinctiveness of direct causation. *Psychonomic Bulletin & Review*.

Young, M. E., & Nguyen, N. (2009). The problem of delayed causation in a video game: Constant, varied, and filled delays. *Learning and Motivation, 40,* 298-312.

Nguyen, N. (2009). *The effect of number of options on choices involving delayed causation*. Unpublished master's thesis, Southern Illinois University, Carbondale, IL.